Discrimination and Goodness of Fit of Multiresponse Mechanistic Models

W. E. Stewart and Y. Shon
Dept. of Chemical Engineering, University of Wisconsin, Madison, WI 53706

G. E. P. Box
Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI 53706

Criteria are derived for objective choice among rival models of multiresponse processes. Formulas are given for the relative posterior probabilities of candidate models and for their goodness of fit as tested on a common data set with Normally distributed errors. The formulas are demonstrated with examples from chemical kinetics and catalysis.

Introduction

Process investigations often generate various mechanistic models and a set of data for testing them. Criteria to identify a preferred model and assess its adequacy are desired. Stewart, Henson, and Box (1996) addressed this problem for models of a single response; in this article we give a corresponding treatment of multiresponse systems.

Our analysis focuses on the following questions: First, which model is most probable according to the data? Second, do any of the models represent the data adequately?

Question one seeks the best approximation to truth, whereas question two is needed to judge whether the model thus chosen is good enough.

Alternative models may be formed from a candidate model by adding or deleting parameters. Additions are favored when they yield higher posterior probabilities; deletions are favored when they do not make the posterior probability appreciably worse. The latter rule adheres to the principle of parsimony, also known as Occam's Razor (Hill, 1977), which states that "multiplicity ought not to be posited without necessity."

Model Discrimination

Consider a table of observations y_{iu} obtained for responses $i = 1, \ldots, m$ in events $u = 1, \ldots, n$ of a process investigation and corrected for any known sources of systematic error. Let $\boldsymbol{\xi}_u$ denote the setting of the vector $\boldsymbol{\xi}$ of independent variables (temperature, pressure, initial concentrations, etc.) in the uth event. Then each observation y_{iu} consists of an expectation value $E(y_{iu}|\boldsymbol{\xi}_u)$ plus an error ϵ_{iu}. Postulating an expectation model M_j with response functions $\{f_{j1}, \ldots, f_{jm}\}$, we get the observation model

$$y_{iu} = f_{ji}(\boldsymbol{\xi}_u, \boldsymbol{\theta}_j) + \epsilon_{iu} \quad \begin{cases} u = 1, \ldots, n \\ i = 1, \ldots, m. \end{cases} \tag{1}$$

For existing observations y_{iu}, this equation gives the errors ϵ_{iu} as functions of the parameter vector $\boldsymbol{\theta}_j$, on the postulate that the model M_j is true. For prospective observations, considered in Eq. 7, y_{iu} will be modeled by the same equation with $\boldsymbol{\theta}_j$ fixed and with errors ϵ_{iu} sampled from an m-variate Normal distribution.

The precision of observations often varies with experimental conditions; therefore, we weight Eq. 1 as follows. Let w_{iu} be the assigned precision (inverse variance) of y_{iu} relative to a standard observation; then multiplication of Eq. 1 by $\sqrt{w_{iu}}$ is appropriate and gives

$$Y_{iu} = \mathcal{F}_{ji}(\boldsymbol{\xi}_u, \boldsymbol{\theta}_j) + \mathcal{E}_{iu} \quad \begin{cases} u = 1, \ldots, n \\ i = 1, \ldots, m \end{cases} \tag{2}$$

for the weighted observations $Y_{iu} := y_{iu}\sqrt{w_{iu}}$ in terms of the weighted expectation functions $\mathcal{F}_{ji} = f_{ji}\sqrt{w_{iu}}$ and weighted errors $\mathcal{E}_{iu} = \epsilon_{iu}\sqrt{w_{iu}}$. Vectors $\boldsymbol{\mathcal{E}}_u = \{\mathcal{E}_{1u}, \ldots, \mathcal{E}_{mu}\}^T$ of weighted errors for prospective events are then modeled as independent samples from an m-variate Normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ with a positive definite covariance matrix $\boldsymbol{\Sigma}$. The ele-

Correspondence concerning this article should be addressed to W. E. Stewart.

ments of Σ are the expectations $\sigma_{ij} = E\{\mathcal{E}_{iu}\mathcal{E}_{ju}\} = \sigma_{ji}$ over this distribution. When Σ is unknown, its sample values will be rendered positive definite by choosing response variables with linearly independent errors over the data; our procedure for doing this is described under "Computer Implementation."

Now consider a set of candidate expectation models, (M_1, ..., M_J). Prior probabilities $p(M_1)$, ..., $p(M_J)$, normally equal, are assigned to the models without reference to data. These values need not sum to 1, since only their ratios affect the ranking of models.

If the models were completely specified, their ranking on the data could be obtained directly via Bayes' theorem (Bayes, 1763; Box and Tiao, 1973). However, each model M_j considered here contains its own unknown parameter vector $\boldsymbol{\theta}_j$, for which a prior probability density $p(\boldsymbol{\theta}_j|M_j)$ is required. We evaluate these densities here by a shortened version of the method of Stewart et al. (1996).

Case 1: Full data, known covariance matrix

For a given experimental design and expectation model, let $p(\boldsymbol{Y}|M_j, \Sigma)$ denote the probability density of prospective data in $\boldsymbol{Y}$-space, predicted by Eq. 2 with an error distribution $N(\boldsymbol{0}, \Sigma)$ for each independent event. Then, according to Bayes' theorem, the *posterior* probability of model M_j conditional on *given* matrices $\boldsymbol{Y}$ and Σ is

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j)\, p(\boldsymbol{Y}|M_j, \Sigma)/C \tag{3}$$

in which C is a normalization constant that is equal for all the models.

If model M_j contains an unknown, estimable parameter vector $\boldsymbol{\theta}_j$, then integration of Eq. 3 over the range of $\boldsymbol{\theta}_j$ gives

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j) \int p(\boldsymbol{Y}|\boldsymbol{\theta}_j, M_j, \Sigma)\, p(\boldsymbol{\theta}_j|M_j)\, d\boldsymbol{\theta}_j/C. \tag{4}$$

This relation can be expressed more naturally as

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j) \int \ell(\boldsymbol{\theta}_j|\boldsymbol{Y}, M_j, \Sigma)\, p(\boldsymbol{\theta}_j|M_j)\, d\boldsymbol{\theta}_j/C \tag{5}$$

by rewriting $p(\boldsymbol{Y}|\boldsymbol{\theta}_j, M_j, \Sigma)$ as a likelihood function for $\boldsymbol{\theta}_j$ conditional on the given $\boldsymbol{Y}$, M_j, and Σ. In this article, the integrations over $\boldsymbol{\theta}_j$ include only those parameters for which unconstrained estimates are obtained.

We treat the prior density $p(\boldsymbol{\theta}_j|M_j)$ for each expectation model M_j as uniform in $\boldsymbol{\theta}_j$ over the region of appreciable likelihood. With this approximation, Eq. 5 gives

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j)\, p(\boldsymbol{\theta}_j|M_j) \int \ell(\boldsymbol{\theta}_j|\boldsymbol{Y}, M_j, \Sigma)\, d\boldsymbol{\theta}_j/C \tag{6}$$

as the posterior probability of model M_j.

The posterior probability $p(M_j|\boldsymbol{Y}, \Sigma)$, being based on data, has a sampling distribution over conceptual replications of the experimental design. We evaluate the ratio $p(\boldsymbol{\theta}_j|M_j)/C$ by requiring that this distribution have expectation $p(M_j)$ if model M_j is true. With this sensible requirement, Eq. 6 yields a posterior probability of the form

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j) \frac{\int \ell(\boldsymbol{\theta}_j|\boldsymbol{Y}, M_j, \Sigma)\, d\boldsymbol{\theta}_j}{E\left[\int \ell(\boldsymbol{\theta}_j|\boldsymbol{Y}, M_j, \Sigma)\, d\boldsymbol{\theta}_j\right]} \tag{7}$$

for each candidate model, with the expectation calculated as if that model were true.

The likelihood function in Eq. 7 for n independent events, each with m responses and the same nonsingular covariance matrix Σ, is

$$\ell(\boldsymbol{\theta}_j|\boldsymbol{Y}, M_j, \Sigma) = |2\pi\Sigma|^{-n/2} \times \exp\left\{-\frac{1}{2}\sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{m}\left[Y_{iu} - \mathfrak{F}_{iu}(\boldsymbol{\theta}_j)\right]\sigma^{ik}\left[Y_{ku} - \mathfrak{F}_{ku}(\boldsymbol{\theta}_j)\right]\right\} \propto \exp\{-S_j(\boldsymbol{\theta}_j)/2\} \tag{8}$$

in agreement with Box and Draper (1965); here σ^{ik} is the (i,k) element of Σ^{-1}. Expressing $S_j(\boldsymbol{\theta}_j)$ as its minimum value $\hat{S}_j$ plus a nonnegative function $G_j(\boldsymbol{\theta}_j)$ and inserting the result in Eq. 7, we obtain

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j) \frac{\exp\{-\hat{S}_j/2\}\int \exp\{-G_j(\boldsymbol{\theta}_j)/2\}\, d\boldsymbol{\theta}_j}{E\left[\exp\{-\hat{S}_j/2\}\int \exp\{-G_j(\boldsymbol{\theta}_j)/2\}\, d\boldsymbol{\theta}_j\right]} = p(M_j) \frac{\exp\{-\hat{S}_j/2\}}{E\left[\exp\{-\hat{S}_j/2\}\right]}. \tag{9}$$

Here we have taken $\exp\{-G_j(\boldsymbol{\theta}_j)\}$ at its expected value for each point in the range of integration; consequently the integrals over $\boldsymbol{\theta}_j$ have cancelled out.

The expectation in Eq. 9 remains to be evaluated on the hypothesis that the model M_j is true. In view of Eq. 8, the minimized sum $\hat{S}_j$ is the sum of squares of the vectors $[\Sigma^{-1/2}\boldsymbol{\mathcal{E}}_u(\hat{\boldsymbol{\theta}}_j)]$, and its sampling distribution for a true model is that of χ^2 with $\nu_j = nm - p_j$ degrees of freedom, where p_j is the number of independent parameters estimated in model M_j. Accordingly, the expectation in Eq. 9 is

$$E\left[\exp\{-\hat{S}_j/2\}\right] = \int_0^\infty \exp(-\chi^2/2)\, p(\chi^2|\nu_j)\, d\chi^2 = \int_0^\infty \exp(-\chi^2)\frac{(\chi^2)^{(\nu-2)/2}}{\Gamma(\nu_j/2)2^{\nu_j/2}}\, d\chi^2 = 2^{-\nu_j/2} \quad \text{with} \quad \nu_j = nm - p_j, \tag{10}$$

and the posterior probabilities in Eq. 7 are

$$p(M_j|\boldsymbol{Y}, \Sigma) = p(M_j)\, 2^{\nu_j/2}\exp(-\hat{S}_j/2) \propto p(M_j)\, 2^{-p_j/2}\exp(-\hat{S}_j/2) \quad j = 1, \ldots, J. \tag{11}$$

This result includes Eq. 12 of Stewart et al. (1996) without the local linearization used there. Normalization of these probabilities to a total of 1 gives

$$\pi(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}) = p(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}) \Big/ \sum_k p(M_k|\boldsymbol{Y},\boldsymbol{\Sigma}) \qquad (12)$$

as the posterior probability share held by model M_j in the candidate set.

The factor $2^{-p_j/2}$ in Eq. 11 is an *a priori* penalty for the number of parameters estimated in model M_j. This penalty offsets the overfitting that occurs when parameters are added to an already adequate model.

Case 1d: Irregular data with known covariances

Irregular data structures, consisting of observations of various subsets of responses, can be handled readily if the relevant covariances are known. The likelihood function is then

$$\ell(\boldsymbol{\theta}_j|\boldsymbol{Y},M_j,\boldsymbol{\Sigma}) \propto \times \exp\left\{-\frac{1}{2}\sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{m}\left[Y_{iu}-\mathfrak{F}_{iu}(\boldsymbol{\theta}_j)\right]\sigma_u^{ik}\left[Y_{ku}-\mathfrak{F}_{ku}(\boldsymbol{\theta}_j)\right]\right\}$$

$$\propto \exp\{-S_j(\boldsymbol{\theta}_j)/2\} \qquad (8a)$$

as shown by Stewart and Sørensen (1981); the resulting function $S_j(\boldsymbol{\theta}_j)$ is a generalization of the one in Eq. 8, with elements σ_u^{ik} from the submatrix $\boldsymbol{\Sigma}_u^{-1}$ computed by inverting the covariance submatrix $\boldsymbol{\Sigma}_u$ of event u. Each $\boldsymbol{\Sigma}_u$ is rendered nonsingular by a proper selection of working response variables; it includes the elements of $\boldsymbol{\Sigma}$ for m_u working responses in the uth event, and placeholder elements δ_{ik} for any responses unused because of their absence or linear dependence. Correspondingly, Y_{iu} and $\mathfrak{F}_{iu}$ are set to zero for any response that is unreported or unused in the uth event.

Equations 9 through 12 follow also from Eq. 8a, with the degrees of freedom revised to $\nu_j = \Sigma_u m_u - p_j$ in Eq. 10. These results are used in Example 1.

Case 2: Full data, full unknown covariance matrix

When $\boldsymbol{Y}$ is full and $\boldsymbol{\Sigma}$ unknown, Eq. 11 takes the form

$$p(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}) \propto p(M_j)2^{-p_j/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{v}(\hat{\boldsymbol{\theta}}_j)\right]\right\} \qquad j=1,\ \ldots,\ J \quad (13)$$

for each of the candidate models. Here $\hat{\boldsymbol{\theta}}_j$ is the maximum-likelihood estimate of the parameter vector $\boldsymbol{\theta}_j$, computable for unknown $\boldsymbol{\Sigma}$ as the minimum point of the determinant function $|\boldsymbol{v}(\boldsymbol{\theta}_j)|$ with elements

$$v_{ik}(\boldsymbol{\theta}_j) = \sum_{u=1}^{n}\left[Y_{iu}-\mathfrak{F}_{ji}(\boldsymbol{\xi}_u,\boldsymbol{\theta}_j)\right]\left[Y_{uk}-\mathfrak{F}_{jk}(\boldsymbol{\xi}_u,\boldsymbol{\theta}_j)\right] \qquad i,k=1,\ \ldots,\ m \quad (14)$$

as shown by Box and Draper (1965).

Treating $\boldsymbol{\Sigma}$ as unknown alters the residual degrees of freedom because of structural differences between Eqs. 11 and 13. The quadratic form $\hat{S}_j$ in Eq. 11 vanishes only if all the residuals do; thus the maximum number of parameters that can be estimated (with a good experimental design) is nm. The determinant $|\boldsymbol{v}(\boldsymbol{\theta}_j)|$ vanishes whenever a linear combination of its rows does; the resulting maximum number of parameters estimable (again with a suitable experimental design) is $n-m+1$. Thus, the minimization of $|\boldsymbol{v}(\boldsymbol{\theta}_j)|$ at rank m leaves $n-p_j$ degrees of freedom for n events, whereas the minimization of $S_j(\boldsymbol{\theta}_j)$ of Eq. 8 leaves $nm-p_j$ degrees of freedom. In compensation for the reduced degrees of freedom, minimization of $|\boldsymbol{v}(\boldsymbol{\theta}_j)|$ gives modal estimates of the corresponding elements of $\boldsymbol{\Sigma}$ as shown by Stewart et al. (1992).

Suppose that $\boldsymbol{Y}$ yields an estimate $\boldsymbol{v}_e$ (from replicates or otherwise) with ν_e degrees of freedom for the experimental error contribution to each matrix $\hat{\boldsymbol{v}}_j := \boldsymbol{v}(\hat{\boldsymbol{\theta}}_j)$. (Here ν_e is the number of replicate events minus the number of distinct settings $\boldsymbol{\xi}_u$ at which replications were done.) Our software selects working responses automatically from those proposed (see "Computer Implementation"), using a smaller working set when necessary so that both $\hat{\boldsymbol{v}}_j$ and $\boldsymbol{v}_e$ are nonsingular and well-conditioned. Subtraction of the constant $\boldsymbol{v}_e$ from $\hat{\boldsymbol{v}}(\boldsymbol{\theta}_j)$ in each of Eqs. 13 gives the proportional expressions

$$p(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}) = p(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}^{-1})$$

$$\propto p(M_j)2^{-p_j/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{v}}_j-\boldsymbol{v}_e)\right]\right\} \qquad j=1,\ \ldots,\ J. \quad (15)$$

The precision matrix $\boldsymbol{\Sigma}^{-1}$ has the conditional probability density

$$p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{Y}) \propto |\boldsymbol{\Sigma}^{-1}|^{(\nu_e-m-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{v}_e\right]\right\} \quad (16)$$

obtainable by Bayes' theorem with a noninformative prior (Jeffreys, 1961; Box and Tiao, 1973) for $\boldsymbol{\Sigma}^{-1}$. With this result and Eq. 8.2.21 of Box and Tiao (1973), Eq. 15 can be averaged over the range of $\boldsymbol{\Sigma}^{-1}$ to obtain the marginal posterior probabilities

$$p(M_j|\boldsymbol{Y}) \propto \int p(M_j|\boldsymbol{Y},\boldsymbol{\Sigma}^{-1})p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{Y})\,d\boldsymbol{\Sigma}^{-1}$$

$$\propto p(M_j)2^{-p_j/2}|\hat{\boldsymbol{v}}_j|^{-\nu_e/2} \qquad j=1,\ \ldots,\ J \quad (17)$$

for the candidate models. Normalization in the manner of Eq. 12 gives the posterior probability share $\pi(M_j|\boldsymbol{Y})$ for each model in the candidate set.

Case 3: Block-diagonal unknown covariance matrix

If the responses can be subdivided into two or more blocks, each having its own independent distribution of error, then the relevant covariance elements form diagonal blocks $\boldsymbol{\Sigma}_b$. The posterior probabilities based on the combined data then take the form

$$p(M_j|\boldsymbol{Y}) \propto p(M_j)2^{-p_j/2}\prod_{b=1}^{B}|\boldsymbol{v}_b(\hat{\boldsymbol{\theta}}_j)|^{-\nu_{eb}/2} \quad j=1, \ldots, J. \tag{18}$$

This result is obtainable by writing $\exp(-\hat{S}_j/2)$ in Eq. 11 as the product of an independent factor for each block, and treating each factor in the manner of Eqs. 13–17. Here the subscript b denotes quantities evaluated for block b, and the elements of $\boldsymbol{v}_b(\boldsymbol{\theta}_j)$ satisfy Eq. 14 for the included response pairs. Normalization to a total of 1 as before gives the posterior probability share for each model in the candidate set.

Three functions are available for computing point estimates $\hat{\boldsymbol{\theta}}_j$ when $\boldsymbol{\Sigma}$ is block-diagonal and unknown. The maximum of Eq. 18 yields estimates $\hat{\boldsymbol{\theta}}_j$ that are optimal for model discrimination, though not the most precise available. More precise parameter estimates are obtained with the function of Box et al. (1972),

$$p(\boldsymbol{\theta}_j|\boldsymbol{Y}) \propto \prod_{b=1}^{B}|\boldsymbol{v}(\boldsymbol{\theta}_j)|^{-n_b/2}, \tag{19}$$

derived by integrating the full posterior density $p(\boldsymbol{\theta}_j, \boldsymbol{\Sigma}^{-1}|\boldsymbol{Y})$ over the permitted range of $\boldsymbol{\Sigma}^{-1}$, or with the function of Stewart et al. (1992),

$$\tilde{p}(\boldsymbol{\theta}_j|\boldsymbol{Y}) \propto \prod_{b=1}^{B}|\boldsymbol{v}(\boldsymbol{\theta}_j)|^{-(n_b+m_b+1)/2} \tag{20}$$

derived by maximizing the full posterior density with respect to $\boldsymbol{\Sigma}$. We prefer Eq. 20 for its superior precision and consistency, and use its point estimates $\hat{\boldsymbol{\theta}}_j$ in Eq. 18 as well.

Goodness of Fit

If the postulated models all fitted the data badly, the preceding results could be misleading. One could be led to choose a poor model because it was *not as bad* as the other models. To guard against such mistakes, scrutiny of the residuals is recommended along with one of the following multivariate tests of goodness of fit.

When $\boldsymbol{\Sigma}$ is known, the goodness of fit of the jth expectation model can be tested by referring the sample value $\chi_j^2 := \hat{S}_j$ in Eq. 8 to the distribution function $P(\chi^2|\nu)$ with $\nu_j = \Sigma_u m_u - p_j$ degrees of freedom. The related function $Q(\chi^2|\nu) := 1 - P(\chi^2|\nu)$, given in Abramowitz and Stegun (1972) and in subroutine form by Press et al. (1992), gives the probability of obtaining a χ^2 value larger than $\hat{S}_j$ from a random replication of the experiments on the hypothesis that the expectation model M_j and the Normal error distribution are true for the experimental situation.

When $\boldsymbol{\Sigma}$ is unknown, but the data permit a sample estimate $\boldsymbol{v}_e$ of the experimental error contribution to $\boldsymbol{v}(\hat{\boldsymbol{\theta}}_j)$, the goodness of fit can be tested via the measure

$$\mathfrak{M}_j := \nu_j \ln|\hat{\boldsymbol{v}}_j/\nu_j| - (\nu_j - \nu_e)\ln|(\hat{\boldsymbol{v}}_j - \boldsymbol{v}_e)/(\nu_j - \nu_e)| - \nu_e \ln|\boldsymbol{v}_e/\nu_e|, \tag{21}$$

in which $\hat{\boldsymbol{v}}_j := \boldsymbol{v}(\hat{\boldsymbol{\theta}}_j)$ is the residual moment matrix with $\nu_j = n - p_j$ degrees of freedom, $\boldsymbol{v}_e$ is the experimental error moment matrix with ν_e degrees of freedom, and $\hat{\boldsymbol{v}}_j - \boldsymbol{v}_e$ is the lack-of-fit moment matrix with $\nu_j - \nu_e$ degrees of freedom. The sampling probability $\Pr(\mathfrak{M} > \mathfrak{M}_j)$ of $\mathfrak{M}$ values exceeding $\mathfrak{M}_j$ is computable from Eq. 26 of Box (1949), under the hypothesis that $\boldsymbol{v}_e$ and $\hat{\boldsymbol{v}}_j - \boldsymbol{v}_e$ are samples from the same multiresponse error distribution. A small value of this probability (say, 0.01 or less) casts doubt on the hypothesis, whereas a value nearer 1.0 supports it. This test, newly installed in our software package GREG, is the multiresponse analog of the two-tailed univariate F-test provided in Box and Tiao (1973).

When $\boldsymbol{\Sigma}$ is block-diagonal, as in Eq. 18, corresponding tests of goodness of fit can be made for the individual blocks of responses once the parameter count for each block is assigned. GREG automates this task, assigning to each block its number p_{jbL} of local parameters plus a fraction $(n'_b - p_{jbL})/(n' - \Sigma_b p_{jbL})$ of the parameters that appear in more than one block. Here n'_b and n' are event counts which include only the first occurrence of each distinct experimental setting. A more detailed assignment may be provided at another time.

This goodness-of-fit test gives a sampling probability by which the adequacy of a model can be judged. Preferably a model should be chosen via Eq. 17 or 18, and then the probability function $\Pr(\mathfrak{M} > \mathfrak{M}_j)$ used as provided in GREG to see if the chosen model is good enough. Strict levels of acceptance or rejection are not advocated because those decisions involve other factors that vary greatly from case to case. Better models would be looked for if the probabilities $\Pr(\mathfrak{M} > \mathfrak{M}_j)$ were all on the order of 0.01 or less, but the decision would depend on the alternatives available and the risks involved in the intended use of such a model.

Computer Implementation

Equations 17, 18, and 21, along with Eq. 26 of Box (1949) for the probability function $\Pr(\mathfrak{M} > \mathfrak{M}_j)$, have been added to the parameter estimation package GREG described by Stewart et al. (1992) for inclusion in a forthcoming book (Stewart et al., 1998). This implementation makes the formulas easy to use. The determinants are evaluated by subroutine DCHDC of Dongarra et al. (1979), modified to accept only those pivotal divisors which remain significant for the corresponding responses. Singularities of the working v-matrices are thus avoided, including the singularity types noted by Box et al. (1973) and also those of McLean et al. (1979).

Examples

Three examples are given here: one on intraparticle diffusion and two on reaction kinetics. Model discrimination with Eqs. 11, 17, and 18 is demonstrated, along with goodness-of-fit testing via Eq. 21 when a suitable error estimate $\boldsymbol{v}_e$ is available.

Example 1: Multicomponent diffusion in a porous catalyst

Feng and Stewart (1974) reported multicomponent diffusion experiments for the system $He-N_2-CH_4$ in an extruded platinum–alumina catalyst over wide ranges of pressure (1–70

Table 1. Discrimination of Models for Diffusion in RD-150 Catalyst

Model	Features	Authors	p_i	$\Sigma_{i,u}\epsilon_{iu}^2$	$\hat{S}_i$	Probability Share $\pi(M_i\|\mathbf{Y})$
1g	Bulk and Knudsen diffusion, one pore-size parameter.	Mason and Evans (1969); Gunn and King (1969)	2	1.655	3,582	10^{-717}
2g	Bulk and Knudsen diffusion, full pore-size distribution $\epsilon(r)$.	Johnson and Stewart (1965)	1	1.431	3,097	10^{-612}
3g	Bulk and Knudsen diffusion, two pore-size parameters.	Feng and Stewart (1974)	4	0.441	954	10^{-147}
3Ag	Bulk and Knudsen diffusion, two pore sizes inferred from $\epsilon(r)$.	Feng and Stewart (1974)	2	0.988	2,138	10^{-404}
4g	Bulk and Knudsen diffusion, one pore-size parameter.	Aris (1965)	2	6.389	13,826	$10^{-2,942}$
1	Model 1g plus two surface diffusivities.	—	4	0.776	1,679	10^{-304}
2	Model 2g plus two surface diffusivities.	—	3	0.717	1,552	10^{-276}
3	Model 3g plus two surface diffusivities.	—	6	**0.128**	**277**	**0.9999**
3A	Model 3Ag plus two surface diffusivities.	—	4	0.204	441	10^{-35}
4	Model 4g plus two surface diffusivities.	-	3	6.239	13,501	$10^{-2,871}$

atm), temperature (300–390K), and composition gradients. The experiments were designed to test several models of diffusion in porous media over the transition range between the free-molecule (Knudsen) and continuum diffusion regimes in a commercial catalyst with a wide pore-size distribution.

The results of the study are summarized in Table 1, along with a brief account of the features of each model. The pore size parameters are regarded here as quadrature abscissae for integration of the flux relations over the pore-size distribution, though some of these authors interpreted their models differently. Each model was fitted by total least squares of 283 observations Y_{iu} of the functions $\ln N_{iuz}$, where N_{iuz} is the measured axial molar flux of species i in the uth event per square centimeter of particle cross-section. In the present notation, that treatment corresponds to setting $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, with a common variance σ^2 for the three responses. Lacking replicates, we compare these models according to Eq. 11 with a variance estimate $s^2 = 0.128/(283-6)$, the residual mean square obtained with model 3. The weighted sum $\hat{S}_j$ in Eq. 11 is then computable for each model as $\Sigma_{i,u}\epsilon_{iuj}^2/s^2 = 2{,}164\Sigma_{i,u}\epsilon_{iu}^2$.

The results thus calculated are striking. Model 3 is overwhelmingly favored, with a posterior probability share exceeding 99.9 percent. This model is also preferable on physical grounds because its physical assumptions are the most realistic. Our variance estimate was chosen by regarding Model 3 as adequate, so the goodness of fit of this model could not be usefully tested. The values of $\chi_j^2 := \hat{S}_j$ for the other models lie outside the range of the tables in Abramowitz et al. (1972), but are evidently large enough for safe rejection.

Example 2: Kinetics of a three-component system

Consider the chemical conversion of initially pure species A_1 to species A_2 and A_3 in an isothermal batch reactor. Simulated yields Y_{iu} at various reaction times, given by Box and Draper (1965), are reproduced in Table 2. Four rival models are postulated for this system:

Model 1. Consecutive irreversible first-order reactions $A_1 \overset{k_1}{\rightarrow} A_2 \overset{k_2}{\rightarrow} A_3$.

Model 2. A reversible first-order reaction $A_1 \underset{k_3}{\overset{k_1}{\rightleftharpoons}} A_2$ followed by an irreversible first-order reaction $A_2 \overset{k_2}{\rightarrow} A_3$.

Model 3. Model 1 supplemented by an irreversible first-order reaction $A_1 \overset{k_3}{\rightarrow} A_3$.

Model 4. Consecutive reversible first-order reactions $A_1 \underset{k_3}{\overset{k_1}{\rightleftharpoons}} A_2 \underset{k_4}{\overset{k_2}{\rightleftharpoons}} A_3$.

Model 1 is the true one, since it was used (along with simulated random errors) to generate the data in Table 2.

Table 3 summarizes our testing of these models against the data of Table 2 using the parameterization $\theta_{jr} = \ln k_r$. The highest posterior probability is found with model 1, and nearly equaled by that of model 3. Models 2 and 4 do not fit well

Table 2. Simulated Data for Example 2

t_u, min	$Y_{1u} = [A_1]$	$Y_{2u} = [A_2]$	$Y_{3u} = [A_3]$
0.5	0.959	0.025	0.028
0.5	0.914	0.061	0.000
1.0	0.855	0.152	0.068
1.0	0.785	0.197	0.096
2.0	0.628	0.130	0.090
2.0	0.617	0.249	0.118
4.0	0.480	0.184	0.374
4.0	0.423	0.298	0.358
8.0	0.166	0.147	0.651
8.0	0.205	0.050	0.684
16.0	0.034	0.000	0.899
16.0	0.054	0.047	0.991

Table 3. Testing of Models for the Data of Table 2

Model	θ_{j1}	θ_{j2}	θ_{j3}	θ_{j4}	p_j	Probability Share, $\pi(M_j\|Y)$	Goodness of Fit, $\Pr(\mathfrak{M} > \mathfrak{M}_j)$
$j=1$	-1.572 ± 0.056	-0.702 ± 0.013			2	0.289	0.91
$j=2$	-1.545 ± 0.11	-0.711 ± 0.14	-4.17 ± 3.6		2	0.216	0.93
$j=3$	-1.632 ± 0.11	-0.800 ± 0.23	-4.31 ± 1.4		3	0.279	0.82
$j=4$	-1.545 ± 0.11	-0.711 ± 0.14	-4.17 ± 3.5	-20.0 $\pm\infty$	3	0.216	0.93

enough to overcome the penalty factor $2^{-1/2}$ for fitting an additional parameter. Sharper discrimination could be achieved by adding experiments of better precision and with other initial compositions, such as pure A_2 and pure A_3. The goodness of fit is acceptable for all four models, and (paradoxically) is best for the obviously overparameterized models 2 and 4.

The simulated data are not mass-balanced. Consequently, the errors in an event do not sum to zero, and GREG retains all three responses in the working set. We prefer to analyze the data as given, rather than normalize them to achieve mass balance, because normalization would remove some covariance information.

Example 3: Kinetics of propylene hydrogenation

The kinetics of propylene hydrogenation over a pulverized platinum–alumina catalyst were investigated by Rogers (1961) and by Shabaker (1965). Their combined data form an array of two rows (responses) and 164 columns (events):

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_2 \end{bmatrix}.$$

The block $\boldsymbol{Y}_1$ consists of the function $\ln \Delta T_{u,\mathrm{obs}}$ for events 1–62 from Shabaker's adiabatic differential reactor. $\boldsymbol{Y}_2$ consists of the function $\ln x_{3u,\mathrm{obs}}$ for events 63–164 from Rogers' 5-bed differential reactor. ΔT_u is a longitudinal temperature difference that is proportional to the reactant conversion, and x_{3u} is the propane production per mole of the reactor inlet stream. Since each response was observed independently, the 2×2 covariance matrix $\boldsymbol{\Sigma}$ is diagonal and Eq. 18 applies.

Kinetic models for this reaction were proposed and tested by Rogers (1961), Shabaker (1965), and Rogers et al. (1966), Mezaki (1968), and Kolboe (1972). These models and many others were compared with the data by Stewart et al. (1988) and with additional models by Lu (1988). The principal models are reexamined here by the present methods.

Table 4 shows the model structures and results. For consistency the data for all temperatures were analyzed simultaneously, as advocated by Blakemore and Hoerl (1963) and by Pritchard and Bacon (1975). Models 1–13 each have two parallel kinetic steps, with the indicated surface species and forward transition complexes all in equilibrium with the adjacent gaseous hydrogen and propylene. Models 1–9 were constructed by Stewart et al. (1988) using plots of the hydrogenation rate versus total reactant pressure, in the manner of Yang and Hougen (1950), but adding more recent evidence for surface species on platinum exposed to gaseous hydrogen or propylene. Models 10, 11, 12, and 13 were proposed by Rogers (1961), Mezaki (1968), Kolboe (1972), and Shabaker (1965), respectively, with less information about the surface species. Models 14 and 15 are included to test a suggestion of Weller (1956) that catalytic reaction rates be modeled as power functions of the partial pressures.

Each model includes a function

$$\alpha_j = \beta_{jd} \exp(\gamma_{jd} t) \quad d = 1, \ldots, 28$$

for the declining catalyst activity $\alpha(t)$ during each of the 28 days of experiments. These functions contribute 55 parameters to each model; the conventional use of graphical activity corrections would implicitly contribute as many parameters or more. With the activity functions included, the experimental precision was assessed by comparing the residuals (for Model 1) of pairs of standard tests normally performed at the beginning and end of the day. We call these pairs "pseudoreplicates" because the quantities compared are constructed with the aid of a model. There were nine such pairs for response 1, and 17 for response 2, giving $\nu_{e1} = 9$ and $\nu_{e2} = 17$. Halving the sum of squared pseudoreplicate residual differences for each response gave the experimental error estimates $v_{e11} = 0.00406$, $v_{e22} = 0.00859$.

The posterior probability shares $\pi(M_j|Y)$ in Table 4 were calculated from Eq. 18; they clearly favor model 1. Stewart et al. (1988) also chose model 1, using a heuristic criterion now superseded by Eq. 18.

The last column of Table 4 shows the goodness of fit of each response by each model. Equation 21 was applied separately to each response, with a parameter count p_{jb} constructed in the manner described under Eq. 21. The probabilities for response 1 with models 1, 2, and 3 are gratifyingly large in view of the approximate nature of the Langmuir adsorption isotherm that they employ. The lower probabilities for response 2 may indicate that the standard tests in that reactor were done with unusual precision, or that the behavior is more complicated at the higher propylene mole fractions investigated in that reactor.

Alternative Discrimination Criteria

The practical importance of model discrimination has motivated many investigations, including criteria and experimental strategies for this purpose. Several known criteria are compared here with ours.

The criteria considered here are as follows:

1. Physical acceptability of the parameters, as in Tschernitz et al. (1946).
2. Compatibility with diagnostic plots of the data, as in Yang and Hougen (1950) and in many textbooks on chemical kinetics.
3. The C_p plot introduced by Mallows (1964, 1973).
4. The information criterion $\mathrm{AIC} = -2(\ln \ell_{j,\max} - p_j)$ proposed by Akaike (1974).
5. Goodness of fit, as in Pritchard and Bacon (1975), Buzzi-Ferraris et al. (1983, 1984, 1990), and Eq. 21 of this paper.

Table 4. Results for Propylene Hydrogenation Models

	Postulated Surface Species																
j	M	H M	C_3H_5 M	C_3H_6 M	C_3H_7 M	$C_3H_6M_2$	$C_3H_5M_3$	H_2 M	Z	H Z	C_3H_5 Z	H_2 Z	Transition Complexes	p_j	Probability Share $\pi(M_j\|Y)$	Goodness of Fit $\Pr(\mathfrak{M} > M_{jh})$ for Y_1 and Y_2
1		+	+	+	+	+	+						C_3H_7M, $C_3H_7M_2$	69	2.1×10^{-1}	1.0×10^{-1} 1.1×10^{-3}
2		+	+		+	+	+						C_3H_7M, $C_3H_7M_2$	67	1.5×10^{-1}	8.3×10^{-2} 7.9×10^{-4}
3		+	+		+	+	+						C_3H_8M, $C_3H_7M_2$	67	1.3×10^{-1}	8.1×10^{-2} 7.6×10^{-4}
4		+	+	+	+	+							C_3H_7M, $C_3H_7M_2$	67	1.0×10^{-1}	2.7×10^{-2} 2.1×10^{-3}
5		+	+	+		+	+						C_3H_7M, $C_3H_7M_2$	67	1.2×10^{-1}	7.7×10^{-2} 7.4×10^{-4}
6		+	+		+	+							C_3H_8M, $C_3H_7M_2$	65	1.2×10^{-1}	2.8×10^{-2} 1.6×10^{-3}
7		+	+		+	+							C_3H_7M, $C_3H_7M_2$	65	1.0×10^{-1}	2.5×10^{-2} 1.7×10^{-3}
8		+	+	+		+							C_3H_7M, $C_3H_7M_2$	65	6.7×10^{-2}	2.3×10^{-2} 1.3×10^{-3}
9		+	+			+							C_3H_6M, $C_3H_7M_2$	63	4.5×10^{-4}	2.7×10^{-3} 1.7×10^{-4}
10	+			+				+	+			+	$C_3H_8M_2$, C_3H_8MZ	63	2.7×10^{-5}	2.6×10^{-4} 1.7×10^{-4}
11	+	+		+					+			+	$C_3H_7M_2$, C_3H_8MZ	65	1.5×10^{-5}	3.2×10^{-3} 1.1×10^{-5}
12	+	+	+						+	+	+		$C_3H_7M_3$, $C_3H_7Z_3$	67	2.1×10^{-7}	2.7×10^{-2} 3.2×10^{-8}
13	+	+				+							$C_3H_7M_3$, $C_3H_8M_2$	63	1.2×10^{-7}	3.0×10^{-4} 1.6×10^{-6}
14	None; two-term power model													63	6.6×10^{-10}	2.4×10^{-8} 1.8×10^{-4}
15	None; one-term power model													59	9.1×10^{-10}	4.0×10^{-8} 7.6×10^{-5}

6. The determinant of the parameter estimation equations, as in Bilardello et al. (1993).

7. Posterior probability as in Box and Hill (1967), with amendments from Box and Henson (1969, 1970), Stewart et al. (1996), and Eqs. 11, 17 and 18 of this article.

Criterion 1 should no longer be used for discrimination because it now can be satisfied directly by including the conditions of acceptability in the parameter estimation. This can be done by using logarithmic parameters to exclude negative rate and equilibrium coefficients, as in Examples 2 and 3, or by using a constrained estimation algorithm as in Stewart et al. (1996).

Criterion 2 is useful for weeding out bad models and devising new ones. This approach is highly recommended.

Criterion 3 was designed for choosing a good predictor from a family of single-response linear regression models. Mallows' statistic C_{pj} for model M_j is an estimated expectation of the scaled sum of squares of prediction errors at the points $\boldsymbol{\xi}_1$, ..., $\boldsymbol{\xi}_n$ and includes a penalty term equal to the number of parameters estimated. This method has worked well for the indicated class of models.

Criterion 4 is a heuristic estimate of the divergence of predictions of a single-response model M_j from those of a comprehensive reference model. The parameterization penalty term proposed by Akaike (1974) is consistent with Mallows (1973). This method has worked well on selection of time-series models, given sufficiently extensive data.

Criteria of goodness of fit (Type 5) are useful for checking the adequacy of candidate models, but are not recommended for discrimination. Such use would tend to favor overparameterized models, as we saw in Example 2. Goodness-of-fit criteria are in fact complementary to posterior probabilities, as noted by Box (1980) in an important paper reconciling frequentist and Bayesian views in applied statistics.

Criterion 6 is useful for comparing experimental designs, as shown by Box and Lucas (1959), but is irrelevant to model discrimination.

Criterion 7, the posterior probability, is the natural and recommended criterion for model discrimination. It yields Eqs. 11, 17, and 18.

Equations 11, 17, and 18 are equally valid whether the experiments are designed beforehand or concurrently. They are preferable to the posterior probability formulas of Box and Hill (1967) for several reasons: They include multiresponse

problems and finite-sample covariance estimates, they exclude prediction covariances from the posterior probabilities, they use fully updated model function values, and they include the penalty factor $2^{-p_j/2}$ discovered by Box and Henson (1969, 1970). These new formulas should lead to improved results in sequential discrimination studies, regardless of the strategy used to choose the settings $\boldsymbol{\xi}_u$. The second and third features listed will prevent inconsistencies like those encountered by Buzzi-Ferraris and Forzatti (1983) and other workers with the procedures of Box and Hill (1967).

Workers on model selection have disagreed regarding the appropriate penalty for the extent of parameterization. Equation 11 applies a penalty factor $2^{-p_j/2}$ to the maximized likelihood function, whereas the formulas of Mallows (1973) and Akaike (1974) give a stronger factor $e^{-p_j/2}$. This difference is to be expected because Mallows sought the model with least squared prediction error, whereas we seek the model with greatest posterior probability. Schwartz (1978) and Chow (1981) obtained a different penalty form by expanding Eq. 6, retaining a term of order $n^{-p_j/2}$ that would factor out if Eq. 7 were used.

Conclusions

Equations 11, 17, and 18 give objective data-based rankings of multiresponse models, as demonstrated in Examples 1, 2, and 3. The single-response discrimination formulas in Eqs. 12 and 17 of Stewart et al. (1996) are directly included. The factor $2^{-p_j/2}$ in these formulas removes the tendency toward overfitting noted by Reilly (1970) and Chow (1981) for discrimination methods based on maximized likelihood.

A goodness-of-fit test for multiresponse models based on the likelihood distribution theory of Box (1949) for the criterion function in Eq. 21 has been given. This test is recommended, along with inspection of residuals, to see if the preferred model represents the data adequately.

Experimental error statistics $\boldsymbol{v}_e$ and ν_e for Eqs. 17, 18, and 21 may be obtained from replicate events or (with abundant data) from residuals of high-order models as in the final example of Stewart et al. (1996). When true replication is prevented by time trends, such as catalyst deactivation, a model of those trends may be used to construct pseudoreplicate residuals as in the present Example 3.

Consistent weighting of the observations and functions is essential in all these calculations. Logarithmic functions were used in Examples 1 and 3 to express the expected uniform relative precision of the raw data. Rearrangement of models into linear forms in the parameters is not recommended, as it makes proper weighting more difficult and is limited to special forms of data, such as isothermal series of differential reactor runs.

Diagnostic data plots greatly expedite model development. Their use should precede and supplement detailed computations on particular models.

Notation

b = response block number, range 1 to B
$\mathrm{E}(Z)$ = expectation (mean) of random variable Z over its distribution
$\boldsymbol{\mathcal{E}}_u$ = column vector of weighted errors in event u
$\mathcal{E}_{iu} = \sqrt{w_{iu}}\,\epsilon_{iu}$, weighted error for response i in event u
$\mathcal{F}_{ji}(\boldsymbol{\xi}_u, \boldsymbol{\theta}_j) = \sqrt{w_{iu}} f_{ji}(\boldsymbol{\xi}_u, \boldsymbol{\theta}_j)$, jth expectation model for weighted observation Y_{iu}
$f_{ji}(\boldsymbol{\xi}, \boldsymbol{\theta}_j)$ = jth expectation model for response i
i, k = response numbers, range 1 to m
j = model number, range 1 to J
ℓ = likelihood function, Eqs. 5 and 8
$\mathfrak{M}_j$ = lack-of-fit statistic, Eq. 21
m = number of working responses (rows) in matrix $\boldsymbol{Y}$
m_u = number of working responses in event u
$N(\mathbf{0}, \boldsymbol{\Sigma})$ = Normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$
$P(\chi^2|\nu)$ = χ^2 distribution with ν degrees of freedom
$p(M_j)$ = prior probability assigned to model M_j
$p(\chi^2|\nu)$ = density of χ^2 distribution with ν degrees of freedom
$Q(\chi^2|\nu) = 1 - P(\chi^2|\nu)$
$S_j(\boldsymbol{\theta}_j)$ = generalized sum-of-squares function in Eq. 8
t = time from start of experimentation on day d, Example 3
$\boldsymbol{v}(\boldsymbol{\theta}_j)$ = matrix function with elements defined in Eq. 14
$\hat{\boldsymbol{v}}_j = \boldsymbol{v}(\hat{\boldsymbol{\theta}}_j)$, maximum-likelihood value of $\boldsymbol{v}(\boldsymbol{\theta}_j)$; see Eq. 20
$\boldsymbol{v}_e$ = experimental error contribution to $\boldsymbol{v}(\boldsymbol{\theta}_j)$
$Y_{iu} = \sqrt{w_{iu}}\, y_{iu}$, weighted observation of response i in event u
y_{iu} = observation of response i in event u

Greek letters

$\Gamma(x)$ = gamma function of x
δ_{ik} = 1 if $i = k$; zero otherwise
ϵ_{iu} = error term in Eq. 1
ν_e = degrees of freedom for experimental error estimate $\boldsymbol{v}_e$
$\boldsymbol{\xi}$ = vector of independent variables
$\boldsymbol{\xi}_u$ = setting of $\boldsymbol{\xi}$ for event u
$\pi(M_j|\boldsymbol{Y}, \bullet)$ = posterior probability share for model M_j
$\prod$ = product sign
$\boldsymbol{\Sigma}$ = covariance matrix for working responses 1, ..., m
$\boldsymbol{\Sigma}_u$ = covariance matrix for working responses in event u
σ^2 = variance of observations of unit weight
χ^2 = coordinate for sums of squares of random samples from univariate Normal distribution $N(0, 1)$

Others

T = transpose of vector or matrix
tr = trace of matrix
$\boldsymbol{A}^{-1}$ = inverse of matrix $\boldsymbol{A}$
$|\boldsymbol{A}|$ = determinant of matrix $\boldsymbol{A}$
$:=$ defines the preceding symbol by the next expression

Literature Cited

Abramowitz, M., and I. A. Stegun, eds., *Handbook of Mathematical Functions*, U.S. Government Printing Office, Washington, DC (1964); reprinted with corrections by Dover, New York (1972).

Aris, R., *Introduction to the Analysis of Chemical Reactors*, Prentice-Hall, Englewood Cliffs, NJ (1965).

Akaike, H., "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, **AC-19**, 716 (1974).

Bayes, T. R., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans. Roy. Soc. London*, **53**, 370 (1763).

Bilardello, P., X. Joulia, J. M. Le Lann, H. Delmas, and B. Koehret, "A General Strategy for Parameter Estimation in Differential-Algebraic Systems," *Comput. Chem. Eng.*, **17**, 517 (1993).

Blakemore, J. W., and A. W. Hoerl, "Fitting Non-Linear Rate Equations to Data," *Chem. Eng. Prog. Symp. Ser.*, **59**, 14 (1963).

Box, G. E. P., "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, **36**, 317 (1949).

Box, G. E. P., "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *J. Roy. Statist. Soc.*, **A143**, 383 (1980).

Box, G. E. P., and N. R. Draper, "The Bayesian Estimation of Common Parameters from Several Responses," *Biometrika*, **52**, 355 (1965).

Box, G. E. P., and T. L. Henson, "Model Fitting and Discrimination," Dept. of Statistics Technical Report 211, Univ. of Wisconsin, Madison (1969).

Box, G. E. P., and T. L. Henson, "Some Aspects of Mathematical Modeling in Chemical Engineering," *Proceedings of the Inaugural*

Conference of the Scientific Computation Centre and the Institute of Statistical Studies and Research, Cairo Univ. Press, Cairo, Egypt, p. 548 (1970).
Box, G. E. P., and W. J. Hill, "Discrimination Among Mechanistic Models," *Technometrics*, **9**, 57 (1967).
Box, G. E. P., W. G. Hunter, J. F. MacGregor, and J. Erjavec, "Some Problems Associated with the Analysis of Multiresponse Data," *Technometrics*, **15**, 33 (1973).
Box, G. E. P., and H. L. Lucas, "Design of Experiments in Non-Linear Situations," *Biometrika*, **46**, 77 (1959).
Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA (1973).
Buzzi-Ferraris, G., and P. Forzatti, "A New Sequential Design Procedure for Discriminating Among Rival Models," *Chem. Eng. Sci.*, **38**, 225 (1983).
Buzzi-Ferraris, G., P. Forzatti, G. Emig, and H. Hofmann, "Sequential Experimental Design for Model Discrimination in the Case of Multiple Responses," *Chem. Eng. Sci.*, **39**, 81 (1984).
Buzzi-Ferraris, G., P. Forzatti, and P. Canu, "An Improved Version of a Sequential Design Criterion of Discriminating Among Rival Response Models," *Chem. Eng. Sci.*, **45**, 477 (1990).
Chow, G. C., "A Comparison of the Information and Posterior Probability Criteria for Model Selection," *J. of Econometrics*, **16**, 21 (1981).
Dongarra, J. J., C. B. Moler, J. R. Bunch, and G. W. Stewart, *LINPACK Users' Guide*, SIAM, Philadelphia (1979).
Feng, C. F., and W. E. Stewart, "Multicomponent Diffusion of Gases in Porous Solids. Models and Experiments," *I&EC Fund.*, **13**, 5 (1974).
Gunn, R. D., and C. J. King, "Mass Transport in Porous Materials under Combined Gradients of Composition and Pressure," *AIChE J.*, **15**, 507 (1969).
Jeffreys, H., *Theory of Probability*, 3rd ed., Clarendon Press, Oxford (1961).
Johnson, M. F. L., and W. E. Stewart, "Pore Structure and Gaseous Diffusion in Solid Catalysts," *J. Catal.*, **4**, 248 (1965).
Kolboe, S., "Kinetics of Olefin Hydrogenation Over a Platinum Catalyst with More Than One Set of Active Sites," *J. Catal.*, **24**, 40 (1972).
Lu, Y. T., "Reaction Modeling of Propylene Hydrogenation over Alumina-Supported Platinum," MS Thesis, Dept. of Chemical Engineering, Univ. of Wisconsin–Madison (1968).
Mallows, C. L., "Choosing Variables in a Linear Regression: A Graphical Aid," Institute of Mathematical Statistics, Manhattan, KS (May 7–9, 1964).
Mallows, C. L., "Some Comments on C_p," *Technometrics*, **15**, 661 (1973).
Mason, E. A., and R. B. Evans, III, "Graham's Laws: Simple Demonstrations of Gases in Motion: I. Theory," *J. Chem. Educ.*, **46**, 358 (1969).
McLean, D. D., D. J. Pritchard, D. W. Bacon, and J. Downie, "Singularities in Multiresponse Modelling," *Technometrics*, **21**, 291 (1979).
Mezaki, R., "Modification of a Reaction Rate Model Involving Competitive-Noncompetitive Adsorption," *J. Catal.*, **10**, 238 (1968).
Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. T. Flannery, *Numerical Recipes in Fortran*, 2nd ed., Cambridge Univ. Press, p. 220 (1992).
Pritchard, D. J., and D. W. Bacon, "Statistical Assessment of Kinetic Models," *Chem. Eng. Sci.*, **30**, 567 (1975).
Reilly, P. M., "Statistical Methods in Model Discrimination," *Can. J. Chem. Eng.*, **48**, 168 (1970).
Rogers, G. B., M. M. Lih, and O. A. Hougen, "Catalytic Hydrogenation of Propylene and Isobutylene over Platinum. Effect of Noncompetitive Adsorption," *AIChE J.*, **12**, 369 (1966).
Rogers, G. B., "Kinetics of the Catalytic Hydrogenation of Propylene," PhD Thesis, Univ. of Wisconsin–Madison (1961).
Schwartz, G., "Estimating the Dimension of a Model," *The Annals of Statistics*, **6**, 461 (1978).
Shabaker, R. H., "Kinetics and Effectiveness Factors for the Hydrogenation of Propylene on a Platinum-Alumine Catalyst," PhD Thesis, University of Wisconsin–Madison (1965).
Stewart, W. E., M. Caracotsios, and J. P. Sørensen, "Parameter Estimation from Multiresponse Data," *AIChE J.*, **38**, 641 (1992).
Stewart, W. E., M. Caracotsios, and J. P. Sørensen, *Computer-Aided Modeling of Reactive Systems*, in preparation (1998).
Stewart, W. E., T. L. Henson, and G. E. P. Box, "Model Discrimination and Criticism with Single-Response Data," *AIChE J.*, **42**, 3055 (1996).
Stewart, W. E., R. H. Shabaker, and Y. T. Lu, "Kinetics of Propylene Hydrogenation on Platinum-Alumina," *Chem. Eng. Sci.*, **43**, 2257 (1988); Errata, **44**, 205 (1989).
Stewart, W. E., and J. P. Sørensen, "Bayesian Estimation of Common Parameters from Multiresponse Data with Missing Observations," *Technometrics*, **23**, 131 (1981); Errata, **24**, 91 (1982).
Tschernitz, J., S. Bornstein, R. B. Beckmann, and O. A. Hougen, "Determination of the Kinetics Mechanism of a Catalytic Reaction," *Trans. Am. Inst. Chem. Engrs.*, **42**, 883 (1946).
Weller, S., "Analysis of Kinetic Data for Heterogeneous Reactions," *AIChE J.*, **2**, 59 (1956).
Yang, K. H., and O. A. Hougen, "Determination of Mechanisms of Catalyzed Gaseous Reactions," *Chem. Eng. Prog.*, **46**, 146 (1950).

Manuscript received Oct. 16, 1997, and revision received Apr. 6, 1998.